








Explainable Multimodal Machine Learning for Engagement Analysis by Continuous Performance Test

Muhammad Arifur Rahman¹ , David J. Brown¹ , Nicholas Shopland¹ ,
Andrew Burton¹ , and Mufti Mahmud^{1,2} 

¹ Interactive Systems Research Group, Computer Science and Informatics,
Nottingham Trent University, Nottingham, UK
{muhammad.rahman02,david.brown}@ntu.ac.uk

² Medical Technologies Innovation Facility, Nottingham Trent University,
Nottingham, UK

<https://www.ntu.ac.uk/research/groups-and-centres/groups/interactive-systems-research-group/>
<https://www.mtif.co.uk/>

Abstract. The human vision system assiduously looks for exciting regions in the real world, in images and videos, to reduce the search effort for various tasks, such as object detection and recognition. A spatial attention representation can divulge the exciting segments, blocks or regions in such images. The Connors' continuous performance test is a visual assessment technique to evaluate the attention and the response inhibition component of executive control to assess attention deficit hyperactivity disorder (ADHD) and other neurological disorders. Artificial Intelligence and Machine Learning models are advancing ever more complex, going from shallow to deep learning over time. Thus, we can achieve higher accuracy and greater precision. However, this also tends to make these models 'black boxes', reducing the comprehensibility of the logic played out in the various predictions and outcomes. This raises an obvious question - how do we understand the prediction suggested or recommended by these machine learning models so that we can place trust in them? XAI attempts to make a trade-off between precision, accuracy and interpretability to achieve this. This research work presents an Explainable Artificial Intelligence (XAI) model for a continuous performance test, monitoring multisensor data and multimodal machine learning for engagement analysis. The sensor data considered included body pose, Electrocardiograph, eye gaze, interaction data and facial features via accurate labelling of engagement or disengagement for cognitive attention of a Seek-X type task execution. We used decision trees and XAI to visualize the multisensor multimodal data, which will help us assess the model's accuracy intuitively and provide us with the explainability of engagement or disengagement for visual interactions.

Keywords: Multimodal learning · Explainable AI · Decision tree · Engagement analysis

1 Introduction

In recent times, inclusive education has been a primary worldwide concern. Researchers across the globe are working towards providing teachers, support staff, and educators with tool-sets to support the assessment and education of children with special educational needs (SEN) as a combined approach towards inclusive education by guiding what pedagogical methodologies are most appropriate for each child depending on their needs. By improving the pedagogical support for these students, they will have an increased chance of inclusion in mainstream classrooms or success in special schools.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects communication and behaviour and can be diagnosed at any stage of life. There is no cure for ASD, but following a diagnosis, early detection of dysregulation events and early intervention may help to diffuse difficult situations in the classroom or at home. With the increasing prevalence of ASD, early detection and possible intervention have become an important challenge [28]. Recently, AI and machine learning (ML) are playing an increasingly dominant role in ASD detection, supporting co-curricular psychology studies. The work of [7, 16, 25, 28] used ML models, performed *in silico* experiments to simplify and assist the conventional clinical experiments in an optimized way.

Besides the SEN students, recently research across the globe has focused significantly on the ability of children with learning difficulties to recognize [6, 33], perceive [31, 38] and interpret [22, 37] emotional cues. So, research on attention detection or recognition of the emotional state of SEN students are still very open. Though ML is used in many cases to develop supportive tools for educators and SEN students, research in this direction and achieving a higher performance is still a challenge. In recent times, artificial intelligence and Machine Learning models are advancing to be ever more complex, going from shallow to deep learning over time. Precisely in this many ML algorithms has been proposed for attention detection i.e. [4, 11, 34, 36] they are considering unimodal data. Until this date, very few machine learning-based methods which consider multiple modalities have been developed for multimodal fusion tasks.

Identification of attention for an individual is challenging and involves multiple factors [8, 48]. Using deep learning models, we can achieve higher accuracy and greater precision. However, this also tends to make these models ‘black boxes’, reducing the comprehensibility of the logic played out in the various predictions and outcomes. This raises an obvious question - how do we understand the prediction suggested or recommended by these machine learning models so that we can place trust in them? Explainable Artificial Intelligence (XAI) [5] attempts to make a trade-off between precision, accuracy and interpretability to achieve this. Here in this work, we presented an XAI ML approach with multimodal data for attention detection.

2 Literature Review

ML became one of the most integrated part in research domain and playing role in many field from genomics analysis [18, 39], image processing [15, 17], text processing [14, 24], trust management [30], different prediction models [26, 41], health care [29, 35] and to a growing list of many more. Even a newer research domain well known as Multimodal Machine Learning (MML) is an emerging multi-disciplinary research domain that enhances the original goals of ML inspired AI by combining multiple complementary and communicative modalities, including vision, text, image, and many more [32].

MML models deal with heterogeneous types of data which bring added challenges to cope with the different modalities, extract data and develop knowledge from it. The process comprises the separate stages of representation, translation, alignment, fusion and co-learning, which is in itself a complex research area. Representation is the study of how to represent and summarise multimodal data which could be complementary or redundant between multiple modalities. The translation is the stage where acquired data is mapped from one modality to another. Due to the heterogeneity of data, this relationship between the modalities is a significant challenge. Alignment is the identification of the relations between multiple modalities. The next step is fusion, where information is joined from multiple modalities to make a prediction, classification or recommendations. Finally, co-learning is the stage where knowledge is transferred between modalities, their representation, and their predictive models [3].

To support practice in academia and various special needs social settings, the demand of AI embedded in non-autonomous systems is gaining interest for human cognition and enhancing learners, support staff and teachers' capabilities. This differs significantly from approaches that aim to create fully automated AI systems. MML and its analytics aim to create AI through externalisation and replicating human cognition and design artefacts closely linked with humans to increase their cognitive abilities and improve their overall capabilities [9, 10, 13].

In a research Hilbert et. al. 2017 used machine learning on multimodal biobehavioral data to classify subjects according to the presence of a generalised anxiety disorder (GAD) from mental disorder (MD) from cortisol data, clinical questionnaire data and structural MRI data using MML [21]. In another study [47]



Fig. 1. The figure shows a Seek-X quiz. Where for a given cue, we need to find the correct answer out of the wrong one. During an experimental setup, the participants were asked to find or seek the target object from different non-target objects acting as a matrix of noise.

used MML for automated international classification of diseases (ICD) coding, where the ICD coding was adopted widely by physicians and other health care workers. Another study by [45] used MML for automatic behaviour analysis to augment clinical resources in diagnosing and treating patients with mental health disorders. In a more recent study, [46] used a multimodal AI-based framework to monitor individual’s working behaviour and stress levels. Identification of this behaviour and stress levels can be achieved with higher precision by fusing multiple modalities obtained from an individual’s behavioural patterns. They used a methodology to determine stress due to workload by integrating heterogeneous sensor data streams, including heart rate, posture, facial expressions and computer interaction.

Early identification can notably improve the prognosis of children with ASD. Yet, existing identification models are expensive, time consuming, and mostly depend on the manual judgment of experts [12,43]. A multimodal framework that can fuse data on a child’s eye movement and facial expression can help identify children with ASD and improve identification efficiency and explainability. Various ML models, used data types and modalities and their performance for attention detection have been summarized in Table 1.

Table 1. Various ML model, used data type and modality for attention detection.

Ref	Data type	ML model	Accuracy	Modality
[4]	EEG, ECG, HF	DWT, SVM	HF 59.64%, EEG 86.86%	Unimodal
[34]	EEG (3 states)	SVM	71.6% to 84.8%	Unimodal
[36]	EEG	SVM	93.33 ± 8.16	Unimodal
[1]	EEG-based passive BCI	SVM	(Avg.) 91.72%	Unimodal
[11]	Eyeball movement	LR, MLP, SVM, DT	LR 96%	Unimodal
[20]	EEG, GSR, ECG,	GR, RF, SVM	Combined 64%	Multimodal
[27]	Facial expression, Eye gaze			Multimodal
[19]	Text and Image	RNN		Multimodal

3 Methodology and Data Sets

3.1 Data Collection

A child’s level of attention can be assessed using mobile devices in a non-intrusive manner. We can observe and record their body posture, facial expression, eye gaze, brain activity (EEG), thermal data, and gesture recognition as forms of data. These data can be collected via different sensors, sometimes wearable and sometimes wirelessly connected. So, a mobile device on which the child is playing a game can be used for a continuous performance test (CPT). The platform tracked students’ engagement, performance and attention with a range of sensors. Head tracking and hand tracking from a RealSense camera combined with head tracking data from a Tobii 4C sensor were used. Body positioning was tracked from the combined posture tracking and gesture tracking data from the

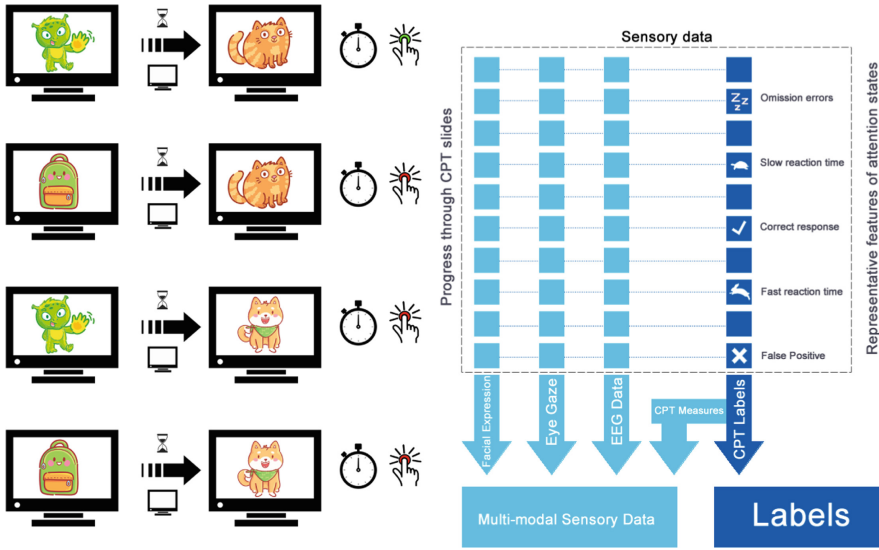


Fig. 2. The figure on the left shows the cartoon of target images that were used in a ‘Where’s Wally’ game. The challenge was to spot Wally in a seek-X type game. The figure on the right shows the multimodal fusion of data obtained from different sensors and their labelling. A detailed explanation is available in [8]

mobile device’s motion sensors. The RealSense camera and the Tobii 4C sensor monitored facial features and eye gaze. A Muse headband (in a child-friendly design) was connected wirelessly over Bluetooth and streamed brain activity data. Figure 2 on the left shows the cartoon of target images that have been used to find- ‘Where’s Wally’ game. Where the challenges to spot Wally, a specified character, a seek-X type games. The figure on the right shows the multimodal fusion of multimodal data obtained from different sensors and their labelling. There were 2615 samples obtained from 59 sessions where 4 participants were involved. An in-detail explanation is available in [8]. Figure 3 shows the basic multimodal data flow evaluation technique. Participants were instructed to find Wally in the seek-X type game. As a part of the CPT experiment, different sensors were collecting multimodal data, such as eye-tracking, facial expression and others. After the labelling of data as by [8], we used our XAI model for attention detection. A detailed explanation of the experimental setup is available at [8].

‘Engagement is the single best predictor of learning in students with learning disabilities’- [23]. In the Swanson’s CPT [44] experiment the participant needs to pay continuous attention to a display screen on an interactive way. Where a game provides them with a pre-defined signal detection challenge. We will say this CPT ‘Seek-X type’ game as [8] to label multi-sensor data. During the experiment, the participants were asked to find the predefined target object from other non-target objects acting as a matrix of noise like a ‘Where’s Wally’ game. The challenge is to spot Wally, from a grid displayed on the screen. The size

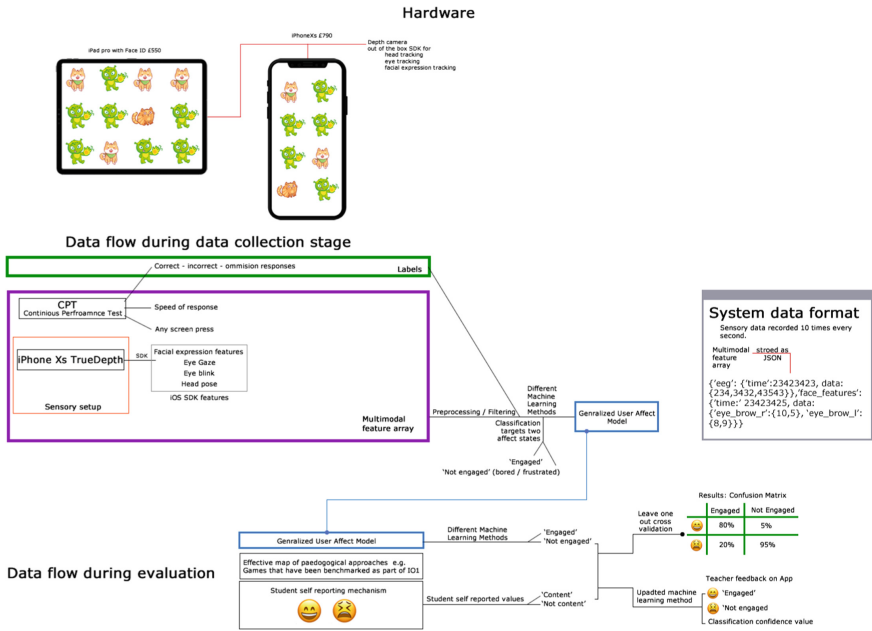


Fig. 3. The figure shows the basic data flow diagram. Participants were instructed to find Wally in the seek-X type game. As part of the CPT experiment, different sensors collected multimodal data like eye tracking, facial expression, and others. After labelling data as by [8], we used our XAI model for attention detection.

of the grids of characters in which to spot Wally in a crowd of characters can be varied. The CPT outcome measures and labels these multimodal data (facial expression, eye gaze, body posture) into high and low attention regions. This provides the labels by which we can assess engagement in the live system.

At the data level, information is highly abstract and the main focus of data fusion is noise reduction and compression. At this level, raw data is processed. Data fusion provides an opportunity for data reduction through data correlations and redundancies. At the feature level, the data has already been processed and the features have been extracted. The fusion is applied to the features themselves rather than the raw data. At the decision level, the data is highly semantic and clear temporal behaviours can be seen in the data. A further detailed explanation of the data preprocessing and fusion is available at [8]. Data frames from these three levels of abstraction with their corresponding CPT attention level labels are used as input into the machine learning layer.

3.2 Decision Tree

The decision tree, a machine learning model, is commonly used in ML, data science and related domains to construct classification tasks based on multiple features or for building prediction algorithms given target variables. If a data

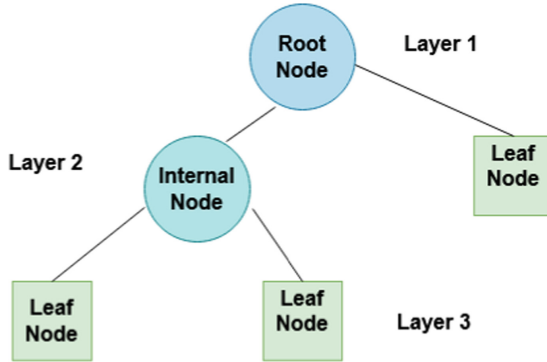


Fig. 4. The figure shows a decision tree where at the root node (layer 1) contains all the instances in a mixed form. Then it splits into two determinations by predictor variable which is also known as a splitting variable that splits between the left child node and the right child node. For a splitting variable, the split criterion depends on some scoring like the Gini Index or Entropy.

set has a mixture of continuous, categorical, and binary types, we can use a decision tree algorithm for better prediction. The decision tree asks yes/no type-specific questions and take decisions. This model classifies a given population into branch-like segments constructing an upside-down tree having multiple levels or heights with the root node on the top level, internal nodes in between levels, and leaf nodes at the bottom. This ML algorithm is a non-parametric model where no parameter tuning is required at the prior stage and can efficiently deal with a large volume of data. The mathematical formulation is also simple and does not impose a complicated parametric structure. Two branches from a parent node are constructed based on the similarity of the data for a given feature, where impurities are calculated by entropy or Gini index. Figure 4 shows a decision tree. During the development of ML models, the data can be divided into two categories. The first segment is the training set, and the second segment is the testing set. A 75% and 25% or 80% and 20% train and test dataset split is a good choice. Yet, k -fold cross-validation is also widely used in the research community for decision trees. However, to leave one out could be a poor choice if the data size is huge. We use the training data set to construct a decision tree and the test dataset to evaluate its performance to construct the final optimal model [40, 42]. We can calculate the accuracy of decision tree algorithm prediction by Eq. 1 where TP indicates the true positive, FP indicates the false positive, TN indicates the true negative, and FN indicates the false negative

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \quad (1)$$

3.3 Gini Index for Decision Tree

Impurities in a decision tree are calculated by the Gini Index (GI), which is also known as Gini impurity. When selected randomly, for a specific feature, GI calculates the probability of that classified incorrectly. If in a single class all the elements or samples are linked with or similar type then this class can be termed as pure. GI varies in the range between 0 and 1, where 0 expresses a pure class obtained from a classification, such that all the elements belong to a specific class, whereas a GI score of 1 indicates an absolute impure or distribution of elements came from a random nature. The GI value is somewhat at the middle shows a nearly equal distribution of samples or elements over some classes. During the modelling phase of the decision tree, the feature providing the least GI value is preferred. The GI can be calculated by Eq. 2 which is determined by calculating the sum of squared probabilities of every classes from one. Mathematically-

$$GI = 1 - \sum_{i=1}^n (P_i)^2 \tag{2}$$

where P_i represents the probability of a sample being classified for a distinct class.

4 Result Analysis

From the figure we can see that the root node starts with 9639 samples of each of the two classes, with a Gini Index. This is a categorical tree where a lower GI represents a better split. Figure 5 and Fig. 6 shows the full splitting mechanism to spilt the data and to measure the decision taking process of attention detection. However, due to the number of levels of the tree, it might not be readable yet a higher resolution image explains the full scenario. To get a better understanding we pruned the number of layers in Fig. 7 just considering four levels. The figure shown in the root node, eye dwelling is the feature that best split the attention and non-attention classes of the data, using as a threshold a value of 22.859.

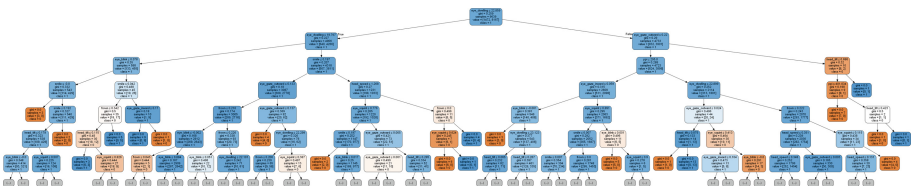


Fig. 5. The figure shows the decision tree for CPT of attention detection from multimodal multi-sensor data up to layer 6. This is an explainable approach and we can easily explain the process of decision making. Due to the size of the tree and the given size of the page, the outcome isn't readable. Yet, a better resolution picture will be easily readable.

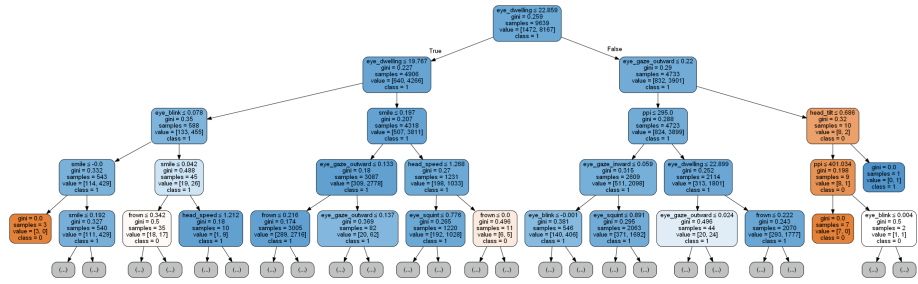


Fig. 6. The figure shows the decision tree for CPT of attention detection from multi-modal multi-sensor data up to layer 4. This is an explainable approach, and we can easily explain the process of decision making.

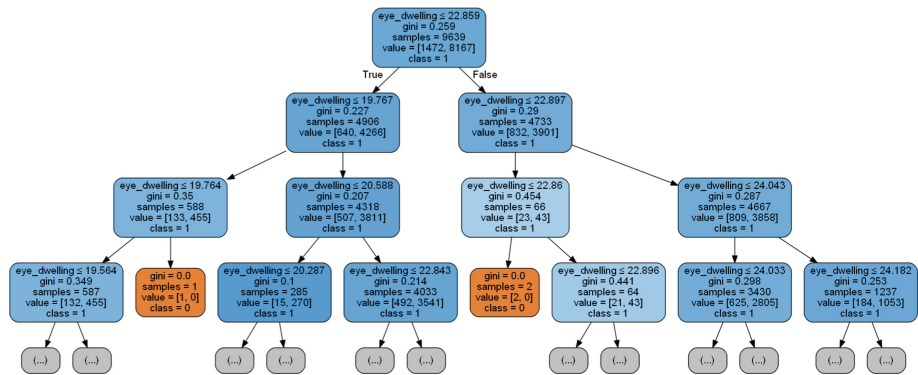


Fig. 7. To get better visualization, we pruned the number of layers in this figure. From the figure we can see that in the root node eye dwelling is the feature that best split the attention and non-attention classes of the data, using as a threshold a value of 22.859. The GI score here is 0.259. From the root node, we get two classes. At level 1, in the left node, we got 4906 samples and in the right node, we got 4733. For the left node of level 1, the threshold value of eye dwelling is 19.767 which splits 4908 samples to further two classes with 640 (left) and 4266 (right) samples with a GI score of 0.227. In the 3rd node of level 2, we can see that the GI score is 0.454, which means both attentive and non-attentive classes are grouped together here.

The GI score here is 0.259 which is not a pure class there are similarities in the chosen class but some impurities are also there. From the root node (let’s say level 0) we get two classes. At level 1, in the left node, we got 4906 samples and in the right node, we got 4733. For the left node of level 1, the threshold value of eye dwelling is 19.767 which splits 4908 samples to further two classes with 640 (left) and 4266 (right) samples with a GI score of 0.227. In the 3rd node of level 2, we can see that the GI score is 0.454, which means both attentive and non-attentive classes are grouped together here. In all of these nodes, all the other features of the data (eye blink, squint, eye gaze inward and outward, facial

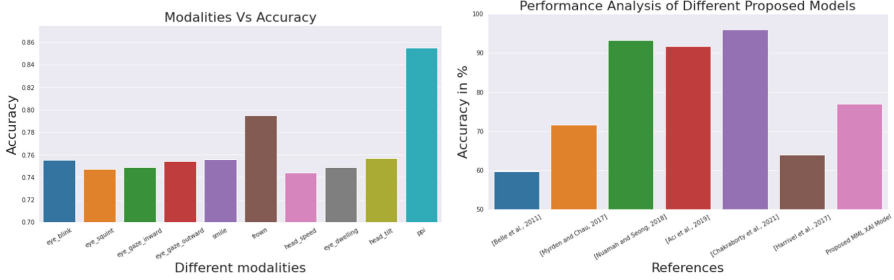


Fig. 8. The figure on the left shows the accuracy for attention detection using an XAI model decision tree for individual modalities. Here we considered eye blink, squint, eye gaze inward and outward, facial feature smile, frown, head tilt and ppi as a unimodal feature. Where the figure on the right shows the comparative performance of our XAI model with different existing ML approaches. The performance of our model is not the best but it did come from an explainable approach. However, as they worked on different dataset the results may also vary as mentioned by [2]

feature smile, frown, head tilt and ppi) were evaluated and had their resulting GI was calculated, however, the decision tree shows that feature that gave us the best results in terms of GI score is eye dwelling.

Figure 8 shows the cooperative performance. In this figure, the left bar graph shows the accuracy for attention detection using our XAI model decision tree for individual modalities. Here the performance for attention detection is shown considering only eye blink, squint, eye gaze inward and outward, facial feature smile, frown, head tilt and ppi as a unimodal feature. Where the right sidebar graph of Fig. 8 shows the comparative performance of our XAI model with different existing ML approaches. The performance of our model is not the best but it did come from an explainable ML algorithm decision tree.

5 Conclusion

In this research work, we presented decision trees from an XAI model for a continuous performance test obtained by monitoring multi-sensor data and multimodal machine learning, for engagement analysis. We considered body pose, eye gaze, interaction data and facial features by objective labelling of engagement or disengagement for cognitive attention of a Seek-X type task execution. We used decision trees, an XAI algorithm, to visualize the decision process of multi-sensor multimodal data, which will help us assess the accuracy of the model intuitively and provide us with the explainability of engagement or disengagement for visual interactions. The accuracy of the model does not give the best possible results, but helps decision making - and it is important that this model is more explainable than the black box-like algorithms of machine learning. As engagement is the single best predictor of learning in students with learning disabilities, we believe, an explainable model for engagement analysis will help

to develop a tool useful in inclusive education by assisting teachers, supporting staff and educators with the assessment of children with SEN.

Acknowledgement. This research was co-funded by the Erasmus+ programme of the European Union in the projects Pathway+ (2017-1-UK01-KA201-036761) ‘A Mobile Pedagogical Assistant to develop meaningful pathways to personalised learning’ and ‘AI-TOP’ (2020-1-UK01-KA201-079167) ‘An AI Tool to Predict Engagement and ‘Meltdown’ Events in Students with Autism’.

References

1. Aci, C.I., Kaya, M., Mishchenko, Y.: Distinguishing mental attention states of humans via an EEG-based passive BCI using machine learning methods. *Expert Syst. Appl.* **134**, 153–166 (2019). <https://doi.org/10.1016/j.eswa.2019.05.057>, <https://www.sciencedirect.com/science/article/pii/S0957417419303926>
2. Adiba, F.I., Islam, T., Kaiser, M.S., Mahmud, M., Rahman, M.A.: Effect of corpora on classification of fake news using Naive Bayes classifier. *Int. J. Autom. Artif. Intell. Mach. Learn.* **1**(1), 80–92 (2020). <https://researchlakejournals.com/index.php/AAIML/article/view/45>
3. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: a survey and taxonomy. *IEEE Trans. Patt. Anal. Mach. Intell.* **41**(2), 423–443 (2019). <https://doi.org/10.1109/TPAMI.2018.2798607>
4. Belle, A., Hobson, R., Najarian, K.: A physiological signal processing system for optimal engagement and attention detection. In: 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW), pp. 555–561 (2011). <https://doi.org/10.1109/BIBMW.2011.6112429>
5. Biswas, M., Kaiser, M.S., Mahmud, M., Al Mamun, S., Hossain, M.S., Rahman, M.A.: An XAI based autism detection: the context behind the detection. In: Mahmud, M., Kaiser, M.S., Vassanelli, S., Dai, Q., Zhong, N. (eds.) *Brain Informatics*, pp. 448–459. Lecture Notes in Computer Science. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-86993-9_40
6. Bloom, E., Heath, N.: Recognition, expression, and understanding facial expressions of emotion in adolescents with nonverbal and general learning disabilities. *J. Learn. Disabil.* **43**(2), 180–192 (2010)
7. Bone, D., Goodwin, M.S., Black, M.P., Lee, C.C., Audhkhasi, K., Narayanan, S.: Applying machine learning to facilitate autism diagnostics: pitfalls and promises. *J. Autism Dev. Disord.* **45**(5), 1121–1136 (2015). <https://doi.org/10.1007/s10803-014-2268-6>
8. Brown, D., Sherkat, N., Taheri, M.: Modeling engagement with multimodal multisensor data: the continuous performance test as an objective tool to track flow. *Int. J. Comput. Inf. Eng.* **14**(162), 197–208 (2020)
9. Brown, D.J., Kerr, S., Wilson, J.R.: Virtual environments in special-needs education. *Commun. ACM* **40**(8), 72–75 (1997). <https://doi.org/10.1145/257874.257891>
10. Brown, D.J., McHugh, D., Standen, P., Evett, L., Shopland, N., Battersby, S.: Designing location-based learning experiences for people with intellectual disabilities and additional sensory impairments. *Comput. Educ.* **56**(1), 11–20 (2011). <https://doi.org/10.1016/j.compedu.2010.04.014>, <https://www.sciencedirect.com/science/article/pii/S0360131510001211>

11. Chakraborty, P., Yousuf, M.A., Rahman, S.: Predicting level of visual focus of human's attention using machine learning approaches. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds.) Proceedings of International Conference on Trends in Computational and Cognitive Engineering, pp. 683–694. Advances in Intelligent Systems and Computing. Springer, Singapore (2021). <https://doi.org/10.1007/978-981-33-4673-4-56>
12. Chen, J., Liao, M., Wang, G., Chen, C.: An intelligent multimodal framework for identifying children with autism spectrum disorder. *Int. J. Appl. Math. Comput. Sci.* (2020). 10.34768/AMCS-2020-0032. <https://sciendo.com/article/10.34768/amcs-2020-0032>
13. Cukurova, M.: Learning analytics as AI extenders in education: multimodal machine learning versus multimodal learning analytics. In: Proceedings of the Artificial Intelligence and Adaptive Education Conference, pp. 1–3 (2019)
14. Das, S., Yasmin, M.R., Arefin, M., Taher, K.A., Uddin, M.N., Rahman, M.A.: Mixed Bangla-English spoken digit classification using convolutional neural network. In: Mahmud, M., Kaiser, M.S., Kasabov, N., Iftekharuddin, K., Zhong, N. (eds.) Applied Intelligence and Informatics, pp. 371–383. Communications in Computer and Information Science. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-82269-9_29
15. Das, T.R., Hasan, S., Sarwar, S.M., Das, J.K., Rahman, M.A.: Facial spoof detection using support vector machine. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds.) Proceedings of International Conference on Trends in Computational and Cognitive Engineering, pp. 615–625. Advances in Intelligent Systems and Computing. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4673-4_50
16. Duda, M., Kosmicki, J.A., Wall, D.P.: Testing the accuracy of an observation-based classifier for rapid detection of autism risk. *Transl. Psychiat.* **4**(8), e424–e424 (2014). <https://doi.org/10.1038/tp.2014.65>, <https://www.nature.com/articles/tp201465>, number: 8 Publisher: Nature Publishing Group
17. Ferdous, H., Siraj, T., Setu, S.J., Anwar, M.M., Rahman, M.A.: Machine learning approach towards satellite image classification. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds.) Proceedings of International Conference on Trends in Computational and Cognitive Engineering, pp. 627–637. Advances in Intelligent Systems and Computing. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4673-4_51
18. Garzon, M., Mainali, S.: Deep structure of DNA for genomic analysis. *Human Molecular Genetics* (2021)
19. Han, Y., Li, L., Zhang, J.: A coordinated representation learning enhanced multimodal machine translation approach with multi-attention, pp. 571–577. Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3372278.3390717>
20. Harrivel, A.R., et al.: Prediction of cognitive states during flight simulation using multimodal psychophysiological sensing–AIAA Information Systems–AIAA Infotech @ Aerospace (2017). <https://doi.org/10.2514/6.2017-1135>, <https://arc.aiaa.org/doi/abs/10.2514/6.2017-1135>, archive Location: world
21. Hilbert, K., Lueken, U., Muehlhan, M., Beesdo-Baum, K.: Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: a multimodal machine learning study. *Brain Behav.* **7**(3), e00633 (2017). <https://doi.org/10.1002/brb3.633>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/brb3.633>

22. Holder, H.B., Kirkpatrick, S.W.: Interpretation of emotion from facial expressions in children with and without learning disabilities. *J. Learn. Disabil.* **24**(3), 170–177 (1991)
23. Iovannone, R., Dunlap, G., Huber, H., Kincaid, D.: Effective educational practices for students with autism spectrum disorders. *Focus Autism Dev. Disabil.* **18**(3), 150–165 (2003). <https://doi.org/10.1177/10883576030180030301>. SAGE Publications Inc Publications Inc Publications Inc
24. Joachims, T.: Learning to classify text using support vector machines, vol. 668. Springer Science and Business Media (2002). https://doi.org/10.1007/978-1-4615-0907-3_2
25. Kosmicki, J., Sochat, V., Duda, M., Wall, D.: Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Transl. Psychiatr.* **5**(2), e514–e514 (2015)
26. Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V., Fotiadis, D.I.: Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **13**, 8–17 (2015)
27. Li, J., Ngai, G., Leong, H.V., Chan, S.C.F.: Multimodal human attention detection for reading from facial expression, eye gaze, and mouse dynamics. *SIGAPP Appl. Comput. Rev.* **16**(3), 37–49 (2016). <https://doi.org/10.1145/3015297.3015301>
28. Liu, W., Zhou, T., Zhang, C., Zou, X., Li, M.: Response to name: a dataset and a multimodal machine learning framework towards autism study. In: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 178–183 (2017). <https://doi.org/10.1109/ACII.2017.8273597>
29. Mahmud, M., Kaiser, M.S., McGinnity, T.M., Hussain, A.: Deep learning in mining biological data. *Cogn. Computat.* **13**(1), 1–33 (2021). <https://doi.org/10.1007/s12559-020-09773-x>
30. Mahmud, M., Kaiser, M.S., Rahman, M.M., Rahman, M.A., Shabut, A., Al-Mamun, S., Hussain, A.: A brain-inspired trust management model to assure security in a cloud based IoT framework for neuroscience applications. *Cogn. Comput.* **10**(5), 864–873 (2018)
31. Metsala, J.L., Galway, T.M., Ishaik, G., Barton, V.E.: Emotion knowledge, emotion regulation, and psychosocial adjustment in children with nonverbal learning disabilities. *Child Neuropsychol.* **23**(5), 609–629 (2017)
32. Morency, L.P., Baltrušaitis, T.: Multimodal machine learning: integrating language, vision and speech. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts, pp. 3–5 (2017)
33. Most, T., Greenbank, A.: Auditory, visual, and auditory-visual perception of emotions by adolescents with and without learning disabilities, and their relationship to social skills. *Learn. Disabil. Res. Pract.* **15**(4), 171–178 (2000)
34. Myrden, A., Chau, T.: A passive EEG-BCI for single-trial detection of changes in mental state. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**(4), 345–356 (2017). <https://doi.org/10.1109/TNSRE.2016.2641956>
35. Nasrin, F., Ahmed, N.I., Rahman, M.A.: Auditory attention state decoding for the quiet and hypothetical environment: a comparison between bLSTM and SVM. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds.) Proceedings of International Conference on Trends in Computational and Cognitive Engineering, pp. 291–301. *Advances in Intelligent Systems and Computing*. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4673-4_23

36. Nuamah, J.K., Seong, Y.: Support vector machine (SVM) classification of cognitive tasks based on electroencephalography (EEG) engagement index. *Brain-Comput. Interf.* **5**(1), 1–12 (2018). <https://doi.org/10.1080/2326263X.2017.1338012>
37. Ouherrou, N., Elhammoumi, O., Benmarrakchi, F., El Kafi, J.: Comparative study on emotions analysis from facial expressions in children with and without learning disabilities in virtual learning environment. *Educ. Inf. Technol.* **24**(2), 1777–1792 (2019). <https://doi.org/10.1007/s10639-018-09852-5>
38. Petti, V.L., Voelker, S.L., Shore, D.L., Hayman-Abello, S.E.: Perception of non-verbal emotion cues by children with nonverbal learning disabilities. *J. Dev. Phys. Disabil.* **15**(1), 23–36 (2003)
39. Rahman, M.A.: *Gaussian Process in Computational Biology: Covariance Functions for Transcriptomics*. Phd, University of Sheffield, February 2018. <https://etheses.whiterose.ac.uk/19460/>
40. Rokach, L., Maimon, O.: Decision trees. In: Maimon, O., Rokach, L. (eds.) *Data mining and knowledge discovery handbook*, pp. 165–192. Springer, US, Boston, MA (2005). <https://doi.org/10.1007/0-387-25465-X-9>
41. Sadik, R., Reza, M.L., Al Noman, A., Al Mamun, S., Kaiser, M.S., Rahman, M.A.: COVID-19 pandemic: a comparative prediction using machine learning. *Int. J. Autom. Artif. Intell. Mach. Learn.* **1**(1), 1–16 (2020)
42. Song, Y.Y., Lu, Y.: Decision tree methods: applications for classification and prediction. *Shanghai Arch. Psychiatr.* **27**(2), 130–135 (2015). <https://doi.org/10.11919/j.issn.1002-0829.215044>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/>
43. Rokach, L., Maimon, O.: Decision trees. In: Maimon, O., Rokach, L. (eds.) *Data Mining and Knowledge Discovery Handbook*, pp. 165–192. Springer US, Boston, MA (2005). <https://doi.org/10.1007/0-387-25465-X-9>
44. Swanson, L.: Vigilance deficit in learning disabled children: a signal detection analysis. *J. Child Psychol. Psychiatr.* **22**(4), 393–399 (1981)
45. Tavabi, L.: Multimodal machine learning for interactive mental health therapy. In: *2019 International Conference on Multimodal Interaction*, pp. 453–456. ICMI 2019. Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3340555.3356095>
46. Walambe, R., Nayak, P., Bhardwaj, A., Kotecha, K.: Employing multimodal machine learning for stress detection. *J. Healthc. Eng.* **2021**, e9356452 (2021). <https://doi.org/10.1155/2021/9356452>, <https://www.hindawi.com/journals/jhe/2021/9356452/>, publisher: Hindawi
47. Xu, K., et al.: Multimodal machine learning for automated ICD coding. In: Doshi-Velez, F., et al. (eds.) *Proceedings of the 4th Machine Learning for Healthcare Conference*. *Proceedings of Machine Learning Research*, vol. 106, pp. 197–215. PMLR, 9–10 Aug 2019. <https://proceedings.mlr.press/v106/xu19a.html>
48. Zubair, M.S., Brown, D.J., Hughes-Roberts, T., Bates, M.: Designing accessible visual programming tools for children with autism spectrum condition. *Universal Access in the Information Society*, pp. 1–20 (2021)